

High-resolution profiling of the LEDGF/p75 chromatin interaction in the ENCODE region

Jan De Rijck¹, Koen Bartholomeeusen¹, Hugo Ceulemans², Zeger Debyser¹ and Rik Gijsbers^{1,*}

¹Laboratory for Molecular Virology and Gene Therapy, KULeuven and IRC KULAK, Kapucijnenvoer 33, B-3000 Leuven and ²Tibotec BVBA, Mechelen, Belgium

Received December 9, 2009; Revised April 23, 2010; Accepted April 30, 2010

ABSTRACT

Lens epithelium-derived growth factor/p75 (LEDGF/p75) is a transcriptional coactivator involved in stress response, autoimmune disease, cancer and HIV replication. A fusion between the nuclear pore protein NUP98 and LEDGF/p75 has been found in human acute and chronic myeloid leukemia and association of LEDGF/p75 with mixed-lineage leukemia (MLL)/menin is critical for leukemic transformation. During lentiviral replication, LEDGF/p75 tethers the pre-integration complex to the host chromatin resulting in a bias of integration into active transcription units (TUs). The consensus function of LEDGF/p75 is tethering of cargos to chromatin. In this regard, we determined the LEDGF/p75 chromatin binding profile. To this purpose, we used DamID technology and focused on the highly annotated ENCODE (Encyclopedia of DNA Elements) regions. LEDGF/p75 primarily binds downstream of the transcription start site of active TUs in agreement with the enrichment of HIV-1 integration sites at these locations. We show that LEDGF/p75 binding is not restricted to stress response elements in the genome, and correlation analysis with more than 200 genomic features revealed an association with active chromatin markers, such as H3 and H4 acetylation, H3K4 monomethylation and RNA polymerase II binding. Interestingly, some associations did not correlate with HIV-1 integration indicating that not all LEDGF/p75 complexes on the chromosome are amenable to HIV-1 integration.

INTRODUCTION

Lens epithelium-derived growth factor p75 (LEDGF/p75) was first described as a transcriptional coactivator and a component of the general RNA polymerase II transcription machinery (1). LEDGF/p75 protects cells against oxidative stress by activation of stress response genes through interaction with stress response elements, and caspase-mediated cleavage of LEDGF/p75 is known to disrupt or reverse this protection (2–6). In addition to mediating stress response, LEDGF/p75 plays a crucial role in cancer (7). Through its interaction with mixed-lineage leukemia (MLL)/menin, LEDGF/p75 is involved in MLL-dependent transcription and leukemic transformation. In agreement, NUP98-LEDGF/p75 fusions were found in human acute and chronic myeloid leukaemia (8–11). LEDGF/p75 is also frequently encountered as an autoantigen in a subset of patients with atopic disorders, mainly atopic dermatitis, and other inflammatory conditions although its function as a potential cause or outcome is not understood (12). It remains to be shown whether the roles of LEDGF/p75 in stress response, cancer and as an autoantigen are interrelated.

A shorter splice variant, p52, is expressed from the same gene as LEDGF/p75 (PSIP1 in humans). While the N-terminal 325 amino acids are identical, LEDGF/p75 has a unique C-terminal region of 205 amino acids. Based on the conserved N-terminal PWWP domain, both isoforms are classified as Hepatoma-derived growth factor-related proteins (HRPs). Together with the AT-hooks and the nuclear localization signal found in the N-terminal part of the protein, the PWWP domain tethers LEDGF/p75 to the chromatin (13,14). The C-terminal moiety of LEDGF/p75 contains the integrase binding domain. Next to binding cellular proteins like JPO2, pogZ, menin/MLL and Cdc7-ASK (7,15–18), this

*To whom correspondence should be addressed. Tel: 32-16-332183; Fax: 32-16-336336; Email: rik.gijsbers@med.kuleuven.be

The authors wish it to be known that, in their opinion, the first two authors should be regarded as joint First Authors.

© The Author(s) 2010. Published by Oxford University Press.

This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (<http://creativecommons.org/licenses/by-nc/2.5>), which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

domain was shown to interact with the integrase protein of human immunodeficiency virus type 1 (HIV-1) and other lentiviruses (19–24). HIV-1 integration into the host cell genome displays a strong bias towards active genes, disfavoring promoter regions and CpG islands (25–28). Knockdown or knockout of LEDGF/p75 abolishes this integration pattern, supporting the hypothesis that LEDGF/p75 binds active genes and tethers the HIV pre-integration complex to these sites via its interaction with integrase (27,29,30). Contradictory to this, cell biology data on LEDGF/p75 predict promoter association (4,6,31,32). To find an answer to this discrepancy, we mapped chromatin binding of LEDGF/p75 by DamID technology and showed that this profile is reminiscent of that of HIV-1 integration. LEDGF/p75 binds active genes disfavoring promoter regions. Correlation with more than 200 genomic features reveals an association with markers of active chromatin. Intriguingly, some markers associate with LEDGF/p75 binding but not with HIV-1 integration suggesting that not all LEDGF/p75 chromatin interactions are amenable to HIV-1 integration.

MATERIALS AND METHODS

Lentiviral vector plasmid construction and vector production

The lentiviral transfer plasmid pLgw-EcoDam-V5-MCS was a kind gift from Dr Bas van Steensel (The Netherlands Cancer Institute, NKI, the Netherlands) and encodes C-terminally V5-tagged *Escherichia coli* Dam methylase (EcoDam) under the control of the HSP (heat shock protein) promoter. The LEDGF/p75 coding region was PCR-amplified using primers F: 5'-*ggg gac aag ttt gta caa aaa agc agg ctt cac tgg cga ttt caa acc tgg*-3' and R: 5'-*ggg gac cac ttt gta caa gaa agc tgg gtc cta gtt atc tag tgt aga atc c*-3' (Invitrogen Gateway[®] recombination region in italics and LEDGF/p75 homology region underlined) and cloned N-terminally to the V5-EcoDam according to the manufacturer's protocol (Invitrogen, Merelbeke, Belgium) yielding plasmid pLgw-EcoDam-V5-LEDGF/p75. The integrity of the construct was verified by DNA sequencing. Lentiviral vectors were prepared as described previously (33). The pLNC LEDGF/p75-IRES-Bsd and pLNC LEDGF/p75D366A-IRES-Bsd plasmids for rescue of HIV replication were described before Ref. (34).

Cell culture and transient transfection

HeLaP4 and 293T cells were grown in Dulbecco's modified Eagle's medium (DMEM) (Gibco-BRL, Merelbeke, Belgium) supplemented with 10% fetal calf serum (Sigma, Bornem, Belgium) and 20 µg/ml gentamicin (Gibco-BRL) (further referred to as DMEM complete) at 37°C and 5% CO₂ in a humidified atmosphere. For western blotting 5 × 10⁵ 293T cells were transfected using 3 µg of plasmid DNA complexed with lipofectamin as described by the manufacturer (Invitrogen) and incubated overnight. HeLa CCR5 A3 cells with a stable LEDGF/p75 knockdown were described before Ref. (34).

Fluorescence laser scanning microscopy

Cells were grown in LabTek II glass chamber slides (VWR International, Haasrode, Belgium) and transfected after 24 h with pmRFP-IN (35) and pLgw-EcoDam-V5-LEDGF/p75 or pLNC LEDGF/p75-IRES-Bsd. The 24-h post-transfection cells were fixed with 4% formaldehyde in PBS for 10 min. Immunohistochemistry staining of endogenous LEDGF/p75 was performed using rabbit anti-LEDGF/p75 antibody (Bethyl, Montgomery, TX, USA). EcoDam-LEDGF/p75 fusion protein was detected using mouse anti-V5 antibody (Sigma). Alexa 488-labelled secondary antibodies were used for detection. Nuclear DNA was stained with 0.5 µg/ml DAPI (Molecular Probes, Merelbeke, Belgium). Confocal microscopy was performed using an LSM 510 meta unit (Zeiss, Zaventem, Belgium). All images were acquired in the multi-track mode. Alexa 488 was excited at 488 nm (by AI laser), mRFP at 543 nm (by HeNe laser) and DAPI at 790 nm (by Spectra-Physics Mai Tai laser). After the main beam splitter (HFT KP 700/543 for mRFP, HFT UV/488/543/633 for eGFP and HFT KP650 for DAPI) the fluorescence signal was divided by a secondary dichroic beam splitter (NFT 490 for eGFP and NFT 545 for mRFP) and detected in the separate channels using the appropriate filters (BP 500–550 for Alexa 488, BP 565–615 for mRFP and BP 435–485 for DAPI).

Western blotting

Protein samples were separated on 10% SDS-PAGE and electroblotted onto polyvinylidene difluoride membranes (PVDF; Bio-Rad). Membranes were blocked with milk powder in PBS/0.1% Tween20 and detection was carried out using mouse anti-V5 antibody (Sigma-Aldrich). Visualization was performed using chemiluminescence (ECL+, Amersham, Diegem, Belgium) using anti-mouse antibodies coupled to HRP (Dako, Denmark).

DamID and DNA tiling array hybridization

Expression of the EcoDam-LEDGF/p75 fusion protein and the EcoDam-only control protein and preparation of gDNA was performed essentially as described (36). Briefly, 5 × 10⁵ HeLaP4 cells were plated in six-well dishes. The next day, three wells each were transduced overnight with 1.5 ml of a lentivector preparation (12.5 × 10³ pg p24/ml) expressing either EcoDam-LEDGF/p75 or EcoDam-only under the control of the HSP promoter or left untreated. After transduction, the medium was replenished and cells were allowed to grow for 48 h under normal cell culture conditions (37°C, 5% CO₂, DMEM complete). After 48 h, cells were collected by trypsinization and gDNA was prepared using the GenElute Mammalian Genomic DNA purification kit (Sigma-Aldrich) according to the manufacturer's protocol. Amplification of methylation-specific PCR fragments was performed as described (36) with minor modifications to comply with the hybridization protocol of the Affymetrix ENCODE 2.0R tiling arrays. For this purpose, PCR amplifications were performed in the presence of dNTP (25 mM each) and dUTP (5 mM). This allowed

for the subsequent digestion of the 200–2000 bp PCR fragments into fragments of ~66 bp using UDG/APE enzymes. The correct length of the fragments was confirmed by electrophoresis using the Agilent Bioanalyzer (Agilent, Diegem, Belgium). Three biological replicates of EcoDam-LEDGF/p75 and EcoDam-only samples each were hybridized with Affymetrix GeneChip ENCODE 2.0R DNA tiling microarrays (Affymetrix, Santa Clara, CA, USA) and subsequently analyzed according to the manufacturer's protocol.

DamID analysis

The MAT (Model-based Analysis of Tiling arrays) algorithm was used to identify genomic regions of statistically significant enrichment of LEDGF/p75 binding (EcoDam-LEDGF/p75 versus EcoDam-only samples) (37). For each array, MAT linearly fits the baseline probe behavior in function of the probe sequence and copy number and uses this model to standardize the readout *t*-value for each probe. A sliding window approach was used to define enriched regions (BandWidth: 300 bp; MaxGap: 300 bp; MinProbe: 10; significance cut-off $P < 10^{-3}$). Other parameters or algorithms (TAS, Tiling Analysis Software; Affymetrix) were also tested and yielded similar results (See also Supplementary Figure S1). The MAT score amounts to the average *t*-value in the probe-centered window and across replicates after trimming of the upper and lower *t*-value deciles, multiplied by the square root of the number of observation points the MAT score is based on. The control MAT score was subtracted from that for LEDGF/p75. Significantly enriched regions were recorded in .bed format for further analysis.

Data analysis

The RefSeq data set and all Encyclopedia of DNA Elements (ENCODE) data sets were obtained from the UCSC website (<http://genome.ucsc.edu/>). The HIV-1 integration set was downloaded from the Bushman Lab website (<http://microb230.med.upenn.edu/>) (28). All data analysis was performed with in-house-written Python scripts (<http://www.python.org/>). To determine the distribution of LEDGF/p75 in transcription units (TUs), each base of a LEDGF/p75 island was given the overall island score. This score was divided into bins according to their relative distance to the transcription start site (TSS). If two TUs overlapped in the opposite strand, the same score was added to both bins. If two or more TUs overlapped in the same strand, the weight of the score was divided by the number of TUs and put in its corresponding bin. An identical approach was followed for the HIV integration sites. In this case, each integration site received a unit weight. To analyze the distribution of LEDGF/p75 around the TSS the same approach was used. In case an LEDGF/p75 positive base was inside one TU and directly upstream of a second TU, the weight of a base was assigned to the former TU since LEDGF/p75 islands do not overlap with TSSs.

Correlation with transcriptional activity

Two publicly available HG-U133A Affymetrix HeLa cell line expression array data sets (GSM156764 from data set GSE9750 and GSM156764 from data set GDS2623) were obtained from the NCBI GEO data set browser (www.ncbi.nlm.nih.gov/geo/). Model-based expression index (MBEI) values were calculated as a measure of the average expression level of individual genes in HeLa cells using the dCHIP algorithm for the combined data sets (38). For each represented gene, the highest expression index was included for further analysis. Genes present in the ENCODE regions were subsequently binned in function of their expression level and the percentage of LEDGF/p75 containing genes was calculated for each bin.

Cross-correlation with ENCODE regions

To calculate the cross-correlation curves between an ENCODE track of interest and the LEDGF/p75 track, the LEDGF/p75 track was shifted along the ENCODE track from -10 kb to +10 kb and the cross-correlation was calculated for each 250 bases (lags) (See Supplementary Figure S3). For cross-correlation with the HIV-1 integration sites, the integration data were blurred to the mean size of the LEDGF/p75 island with scores declining from the island center to the borders. To test for significance, 600 random LEDGF/p75 and integration tracks of equal size as the observed tracks were generated. To generate random LEDGF/p75 islands tracks, islands of the same size and MAT score were picked at a randomly chosen location in the ENCODE region. To generate the matched HIV-1 integration site tracks for each experimental integration site at an observed distance from the cloning restriction site, a control site was selected within the ENCODE regions at the same distance of a randomly chosen restriction site. The 2.5 through 97.5 percentile interval for control Pearson's cross-correlation coefficients with the ENCODE track of interest was calculated at lag 0. Coefficients of random tracks at lower or higher lags did not statistically differ from lag 0 in a few representative trials and were further omitted to minimize processing time.

Cytological analysis of HeLaP4 and Jurkat cells

Genomic DNA from Jurkat, HeLaP4 and from peripheral blood mononuclear cells from a healthy individual were extracted and labeled with Cy5 or Cy3 using the Bioprime Labeling kit (Invitrogen) according to the manufacturer's protocol. Probes were hybridized on Syndrome Plus v2 array chips (Oxford Gene Technology, Yarnton, UK) using the Oligo aCGH hybridization kit (Agilent) according to the manufacturer's protocol. Chips were scanned on an Agilent microarray scanner and analyzed using Agilent Feature Extraction software and Cytosure software (Oxford Gene Technology). Cy5-labeled Jurkat or HeLaP4 DNA was analyzed together with Cy3-labeled control DNA on the same chip. The analysis for each cell line was repeated with switched dyes.

RESULTS

Expression and functional validation of the EcoDam-LEDGF/p75 fusion protein

To determine the LEDGF/p75 chromatin interaction profile we used DamID technology, which relies on the expression of a fusion between the EcoDam and a chromatin binding protein of interest (36,39). Chromatin binding results in methylation of neighboring GATC sites that are specifically amplified. The resulting probes are used for microchip analysis. Expression of the EcoDam-LEDGF/p75 fusion protein is driven by an inducible HSP promoter and encoded in a lentiviral vector (Figure 1A). Stable HeLaP4 cells were generated (referred to as HeLaP4 EcoDam-LEDGF/p75) together with an

EcoDam cell line (HeLaP4 EcoDam) to control for aspecific methylation. Under normal cell culture conditions (37°C), the HSP promoter hardly supports expression of the EcoDam-LEDGF/p75 fusion or EcoDam protein. Low expression minimizes aspecific chromatin binding and limits the effect of the exogenous protein on the cellular environment. Accordingly, expression in these cell lines was undetectable via immunocytochemistry or western analysis (data not shown). However, transient transfection of lentivector transfer plasmids could confirm expression of EcoDam-LEDGF/p75 and EcoDam by western blot analysis (Figure 1B). EcoDam-LEDGF/p75 displayed the typical, dense, fine-speckled subnuclear distribution pattern of wild-type LEDGF/p75 as observed by confocal microscopy (Figure 1C).

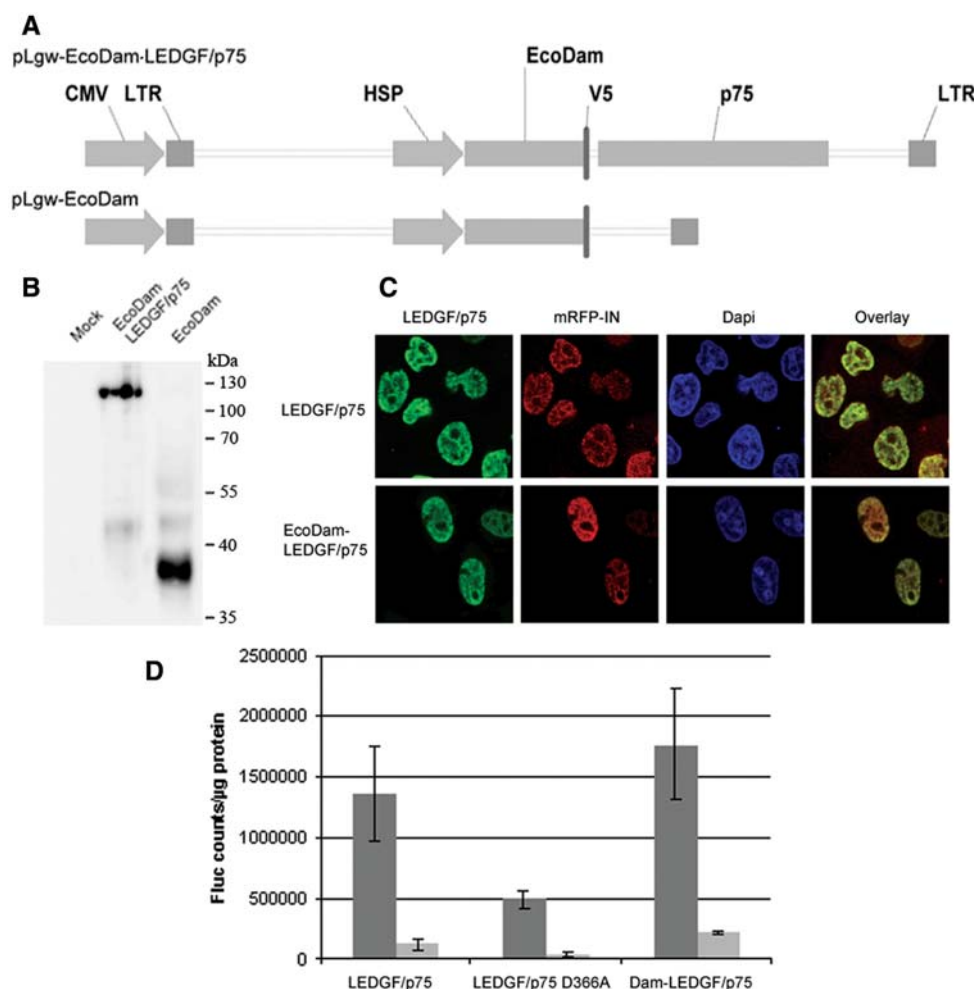


Figure 1. The EcoDam-LEDGF/p75 fusion protein behaves like endogenous LEDGF/p75. **(A)** Schematic representation of the pLgw-EcoDam-V5-LEDGF/p75 transfer plasmid expressing the EcoDam-V5-LEDGF/p75 fusion protein, and the pLgw-EcoDam-V5 transfer plasmid expressing the EcoDam-V5 control protein. CMV, cytomegalovirus promoter; LTR, long terminal repeat; p75, LEDGF/p75. **(B)** 293T cells were transiently transfected with the respective transfer plasmids. Total cell lysate was prepared and analyzed by western blotting using an anti-V5 antibody. **(C)** HeLa CCR5 A3 cells depleted for LEDGF/p75 (34) were transfected with a LEDGF/p75 (upper panels, green) or an EcoDam-V5-LEDGF/p75 expression plasmid (lower panels, green) together with an mRFP-integrase expression plasmid (red). DAPI was used to stain the DNA. An overlay of the green and red panels is shown. **(D)** HIV-1 replication is rescued by the EcoDam-LEDGF/p75 fusion protein. HeLa CCR5 A3 cells were transfected with the indicated constructs. The 24-h post-transfection cells were infected with two dilutions of an HIV-1 clone encoding firefly luciferase (NL4.3 fLuc) (dark and light gray). Luciferase expression was measured 48 h post-infection. Fluc counts upon transfection of a control plasmid encoding an EcoDam fusion to CBX1 (51) were subtracted. LEDGF/p75 D366A, HIV integrase interaction deficient LEDGF/p75 (40). The experiment was done in triplicate and repeated four times. A representative experiment is shown. Error bars indicate the SD of the triplicate values.

Moreover, like wild-type LEDGF/p75, EcoDam-LEDGF/p75 co-localized with HIV-1 integrase (23). Full functionality of the EcoDam-LEDGF/p75 fusion protein was evidenced in an HIV-1 rescue experiment (Figure 1D). HeLa CCR5 A3 cells stably depleted for LEDGF/p75 (34) were back-complemented with the EcoDam-LEDGF/p75 expression construct. As a control, wild-type LEDGF/p75 and a non-interacting LEDGF/p75 mutant (D366A) (40) were used to back-complement the cells. The EcoDam-LEDGF/p75 fusion protein could rescue infection with HIV-1 Fluc to the same extent as wild-type LEDGF/p75.

LEDGF/p75 associates with transcription units, downstream of the transcription start site

LEDGF/p75 chromatin binding regions were determined by hybridization to ENCODE DNA tiling arrays (41). The ENCODE project characterized 44 regions covering about 1% of the human genome, comprising loci of specific interest, such as the HOXA cluster, together with randomly selected regions. More than 200 specific tracks have been published that record distribution of various epigenetic modifications and transcription factor binding sites in these regions, explaining our choice for ENCODE in this study. The correlation of many of these tracks with HIV-1 integration has been studied before (28). We performed and validated three independent DamID experiments in HeLaP4 cells. The MAT algorithm identified 644 LEDGF/p75 chromatin binding islands from the combined data sets and corresponding controls (37). Islands were heterogeneous with regard to their size and LEDGF/p75 binding intensity (MAT score): sizes ranged between 415 and 2941 bases, with an average of 1025.03 and an SD of 361.43 bases; MAT scores varied between 1.82 and 9.89, with a mean score of 3.18 ± 1.05 . One by one comparison of the experiments showed an overlap of $\pm 50\%$ of the islands (Supplementary Table S1). Around 25% of the islands were common to all three experiments and around 40% were unique. Different parameters were tested during MAT analysis to define the LEDGF/p75 islands (See 'Materials and Methods' section and Supplementary Figure S1). Downstream analysis yielded the same results as the LEDGF/p75 islands used in this article.

Random control sites were generated computationally. Per experimental site 10 sites were computed, matched with respect to the size and MAT score at a randomly chosen location in the ENCODE region [matched random control, (MRC)]. In the analyses that follow, the distribution of experimental LEDGF binding sites is compared to that of the MRC sites. The total RefSeq TUs coverage of ENCODE regions is 45.7%, which is about 10% higher than for the whole genome. Seventy-five percent of the LEDGF/p75 islands were found in RefSeq TUs as compared to 47% of the MRC islands (Figure 2A). Significantly, fewer LEDGF/p75 islands (0.3%) than random (1.5%) overlapped with the TSS ($P < 0.05$, χ^2 -test). Conversely, significantly more LEDGF/p75 islands (2.3%) than random (1.3%) coincided with the transcription termination site ($P < 0.05$, χ^2 -test). Similar results were obtained for Ensembl TUs (Figure 2A). Given the

fact that many TUs overlap, we repeated the analysis for the 268 non-overlapping genes in the ENCODE regions; none of the LEDGF/p75 islands comprised a TSS (Figure 2A, non-overlap).

Taken together, these data demonstrate that LEDGF/p75 preferentially associates with TU downstream of the TSS. Interestingly, 86% of the HIV integration sites in the ENCODE region were also located in a TU (28). Significantly, fewer RefSeq TUs were bound by LEDGF/p75 (28.4%) than expected from MRC (37.9%) ($P < 0.0001$, χ^2 -test), suggesting that LEDGF/p75 preferentially targets selected subsets of genes. However, gene ontology analysis did not show significantly favored or disfavored functional GOstat categories in comparison with all ENCODE genes (42). Furthermore, the GIMSAM search engine failed to detect a binding site that was significantly enriched as compared to MRC (43).

To take a closer look at the distribution of LEDGF/p75 binding sites within TUs, LEDGF/p75 islands were binned relative to their position in the TU and the cumulative MAT score for each bin was plotted (Figure 2B). The probability of LEDGF/p75 binding gradually decreases towards the transcription termination site. The downward trend was also observed not taking into account the MAT scores, but only the presence of LEDGF/p75 (data not shown). Linear regression qualified this negative trend as significant ($P < 0.0001$, Fisher's z). Analysis of the mean MAT score per bin or the distribution of island sizes over the TU did not show significant differences ($P > 0.1$, Fisher's z). This trend was not observed for MRC islands too ($P = 0.13$, Fisher's z). As a consequence, the amount of islands is significantly higher at the 5'-site of TUs. Next, we binned MAT scores in function of their absolute distance to the TSS and plotted the MAT score sums for the bins within 10 kb around the TSS. The probability of LEDGF/p75 binding is low upstream of the TSS (Figure 2C), but rises steeply immediately downstream of the TSS. Wherever islands mapped to multiple, partially overlapping genes, we preferentially assigned islands to genes that completely encompass the island, because unequivocally assigned LEDGF/p75 islands almost never include a TSS. As a consequence, the MAT score sums of the MRC islands also increased downstream of the TSS. Importantly, while the profiles are similar upstream of the TSS and near the TSS, the rise of the LEDGF/p75 profile downstream of the TSS is about three times higher than that of MRC, again indicating an enrichment of LEDGF/p75 downstream of the TSS. Since the probability of LEDGF/p75 binding decreases towards the transcription termination site, LEDGF/p75 binding around the transcriptional end was indistinguishable from MRC (Figure 2D). In conclusion, LEDGF/p75 preferentially associates with chromatin downstream of the TSS and binding decreases towards the transcriptional end.

LEDGF/p75 chromatin binding correlates with transcriptional activity of the transcriptional unit bound

HIV-1 integration correlates positively with transcriptional activity of the targeted TU (26). To study a possible

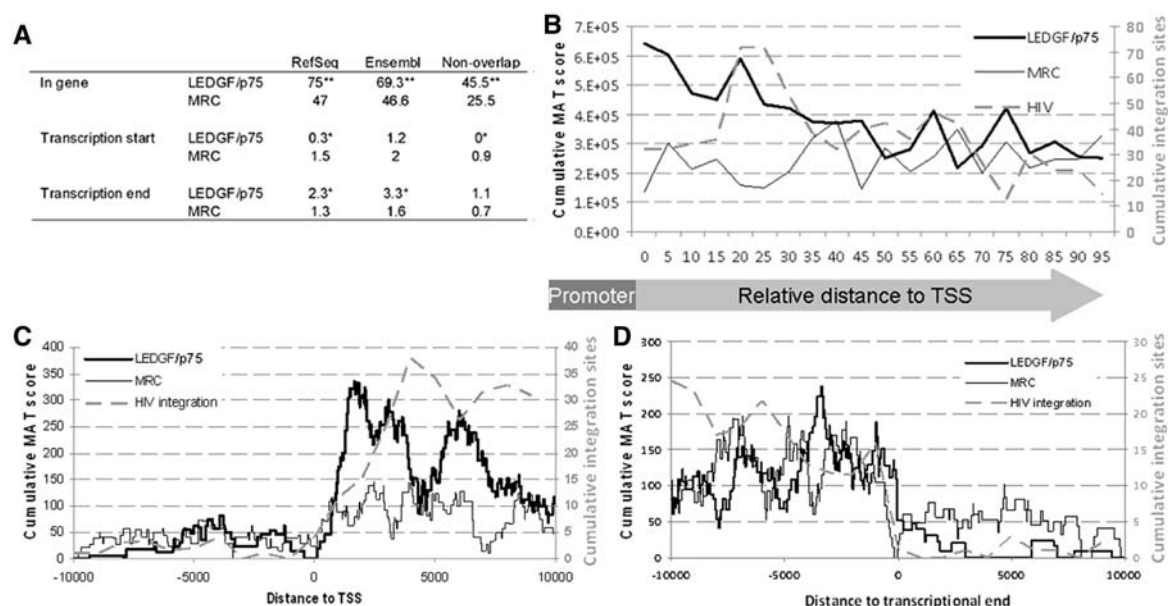


Figure 2. LEDGF/p75 binds TUs downstream of the TSS. (A) Distribution of LEDGF/p75 islands around TUs. The percentage of LEDGF/p75 islands in TUs and overlapping with TSS or transcription ends is shown for RefSeq and Ensembl TUs and for non-overlapping RefSeq TUs. As a control, a random data set of LEDGF/p75 islands was generated 10 times the size of the original data set (MRC). * $P < 0.05$, ** $P < 10^{-27}$ as compared to random as measured by a χ^2 -test. (B) LEDGF/p75 binding in TUs was plotted relative to the TSS (LEDGF/p75). All genes were divided in 20 equal bins and the MAT score of each LEDGF/p75 positive base was assigned to the corresponding bin. For each LEDGF/p75 island, 10 new islands were picked at a random position within the ENCODE region having the same size and score as the parental island (MRC). As for LEDGF/p75 islands, the random islands were plotted relative to the TSS. Values were divided by 10 for comparison with the LEDGF/p75 islands. HIV-1 integration sites in TUs (28) were assigned to the corresponding bin relative to the TSS (HIV integration). If LEDGF/p75 or integration sites could be assigned to more than one TU, the weight was divided over all TUs. (C) Distribution of LEDGF/p75, MRC and HIV-1 integration sites around the TSS. Scores for LEDGF/p75 and random islands were accumulated for each base in a 20kb window around the TSS. All TUs were analyzed in 5'-3' orientation. Data values of the random curve were divided by 10 to be comparable with LEDGF/p75 islands. For HIV-1 integration sites, the 20kb window was divided into 1-kb bins. Each integration site was assigned to the corresponding bin. Since LEDGF/p75 islands almost never overlap with the TSS, a positive base found in the TU was fully assigned to this TU and not divided over potential neighboring units. (D) Distribution of LEDGF/p75, random islands (MRC) and HIV-1 integration sites around the transcriptional end. Analysis was performed as described in C.

association between transcriptional activity (derived from the NCBI GEO data set) and LEDGF/p75 chromatin binding, ENCODE genes were binned in function of their transcriptional activity and the percentage of LEDGF/p75 positive genes for each bin was calculated (Figure 3). LEDGF/p75 binding was shown to correlate significantly with gene activity (multinomial Fisher's exact test, $P < 0.05$), while no correlation was demonstrated for the MRC ($P = 0.8$).

HIV-1 integration is targeted to LEDGF/p75 chromatin binding islands

HIV-1 integration is targeted to active TUs, and knockdown or knockout of LEDGF/p75 has been reported to disrupt this targeting preference (27,29,30), suggesting a direct role for LEDGF/p75 in HIV-1 targeting. In addition, by replacement of the N-terminal chromatin binding domains of LEDGF/p75 by alternative chromatin binding domains, retargeting of viral integration was recently achieved (34,44,45). More than 800 HIV-1 integration sites were identified in ENCODE regions in Jurkat T cells (28). A quantitative model that predicts HIV-1 integration sites based on genomic features determined several ENCODE regions to contain more or less integration sites than expected (28,46). For instance, ENCODE region ENm014 on Chromosome 7 carries

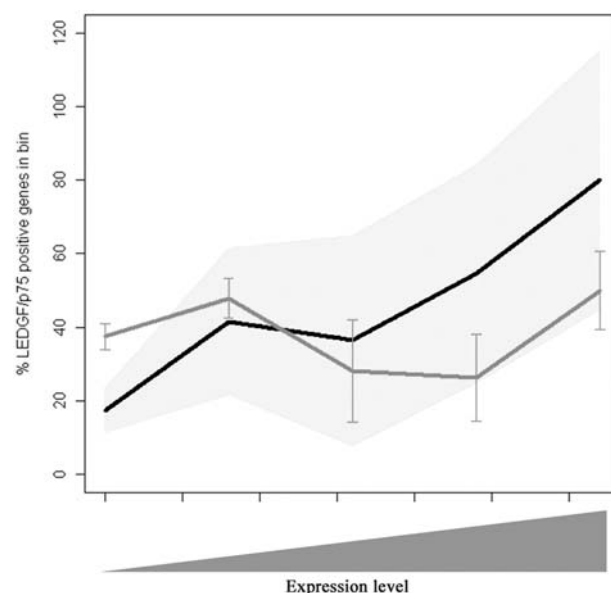


Figure 3. LEDGF/p75 chromatin binding correlates with transcriptional activity. Genes in the ENCODE region were divided into bins according to their expression level. The percentage of LEDGF/p75 positive genes in each bin is shown in black. The 95% confidence interval is shown (gray shading). As a control, 10 MRC sets of equal size as the original data were computed and divided over the bins (gray line). The mean \pm SD is shown.

three integration sites, all located within the same TU (ZNF800; Figure 4A). Strikingly, we also found a major cluster of LEDGF/p75 islands in ZNF800. For a more exhaustive study of the association of HIV-1 integration sites and LEDGF/p75 chromatin binding sites, we generated MRC integration sites (HIV MRC). To this end, for each experimental integration site at an observed distance from the cloning restriction site, 10 control sites were computed within the ENCODE regions each at the same distance of randomly chosen restriction sites. This approach controls for recovery bias due to cleavage by restriction enzymes (46,47). Of 861 HIV-1 integration sites, 4.65% were found in the LEDGF/p75 islands, as compared to only $2.43 \pm 0.45\%$ of the HIV MRC sites ($P < 0.01$, Fisher's exact test). We then assigned HIV integration sites to the nearest LEDGF/p75 island and computed the percentage of integrations within absolute distance bins from the center of the LEDGF/p75 islands (Figure 4B). Zooming in on the center of LEDGF islands (<3.5 kb from the center), HIV-1 integrates three times more frequently near the center of LEDGF/p75 islands than HIV MRC. Beyond 3.5 kb from the island center, the enrichment attenuates and becomes indistinguishable from that of HIV MRC, around 10 kb from the center. Since the HIV integration

data and the LEDGF/p75 chromatin binding data were obtained in two different cell lines, possible biases due to a different expression profile or different chromosomal content of the cells had to be controlled for. To analyze whether different expression levels could influence the results, HeLa and Jurkat expression profiles were derived from the NCBI GEO data set and compared. Chromatin regions with genes showing more than 10-fold difference in expression were omitted from the analysis. Also, for this data set, integration preferentially took place in the neighborhood of LEDGF/p75 islands, ruling out that observed differences were due to different expression profiles (Supplementary Figure S2a). To analyze the status of the ENCODE regions in Jurkat and HeLaP4 cells, genomic DNA was extracted and analyzed on OGT Syndrome Plus v2 array chips. These chips are especially designed to detect genome-wide chromosomal aberrations. The results showed that Jurkat cells are near diploid. Only one of the 44 ENCODE regions contained a small deletion. HeLa cells contain more chromosomal aberrations although at least one copy of all ENCODE regions was present. Fourteen out of 44 ENCODE regions showed duplications or deletions in one allele. As most of these regions seemed to be mosaic, it was impossible to determine an exact copy

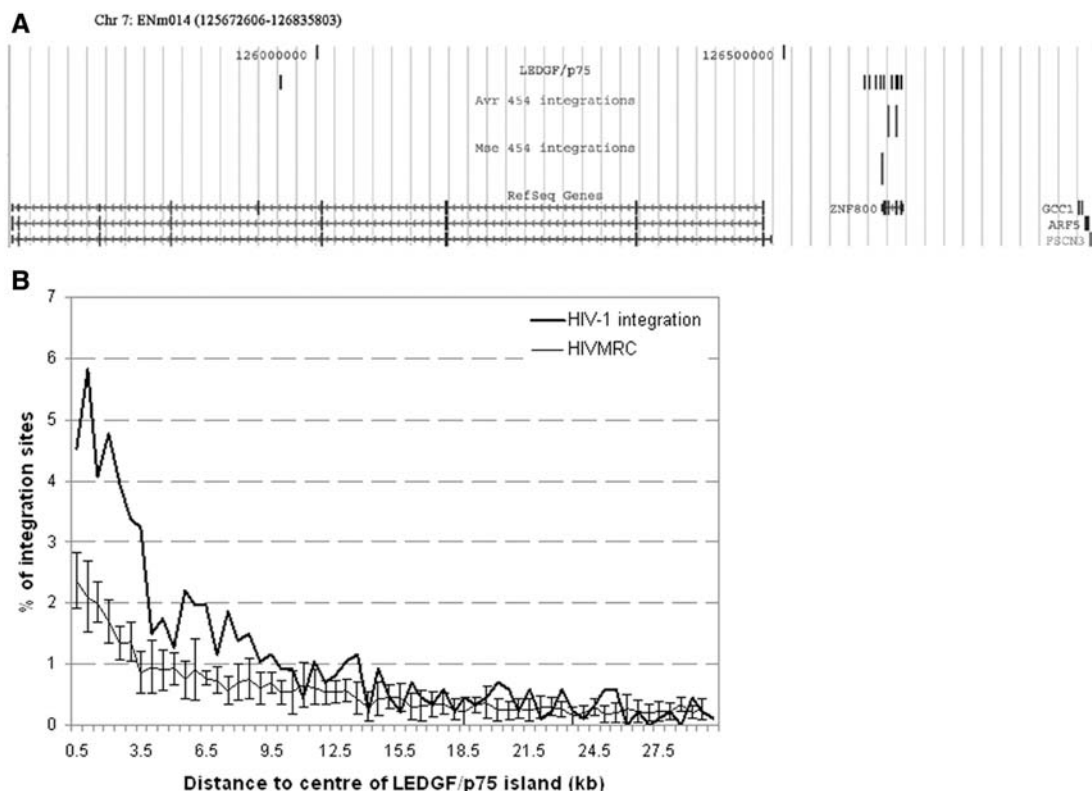


Figure 4. HIV-1 integration is targeted to LEDGF/p75 chromatin binding islands. (A) Schematic representation of LEDGF/p75 islands and HIV-1 integration sites in ENCODE region ENm014 (<http://genome.ucsc.edu/>). This ENCODE region contains significantly less integration sites than predicted (28). All integration events took place around the same TU that also contains LEDGF/p75 islands. GCC1, GRIP and coiled-coil domain containing 1; ARP5, actin-related protein 5 homolog; FSCN3, fascin 3. (B) The region surrounding the middle of LEDGF/p75 islands was divided in 0.5 kb bins. Integration sites were assigned to the closest LEDGF/p75 island and placed into the corresponding bin. The percentage of integration sites in each bin is shown (bold line). Ten data sets of matched random integration sites (HIV MRC) of equal size as the HIV-1 integration data set were generated and assigned to the corresponding bin (light line). The SD is shown.

number. For this reason, we removed all non-diploid ENCODE regions and reanalyzed the correlation between HIV integration sites and LEDGF/p75 binding (Supplementary Figure S2b). Also under these conditions, the association of LEDGF/p75 with genuine HIV integration sites remained tighter than that with MRC integration sites.

Association of LEDGF/p75 with markers of active transcription

The correlation of HIV-1 integration with ENCODE tracks has been studied before (28). As a logical next step, we performed a similar analysis for the LEDGF/p75 chromatin binding sites. We quantified the cross-correlation of the observed LEDGF/p75 binding sites, known HIV-1 integration sites, and 600 MRC or HIV MRC tracks with 215 ENCODE tracks (41). Sixty-two tracks did not correlate differently with LEDGF/p75 binding sites or the MRC; for 126 tracks a significantly positive and for 27 a significantly negative correlation was obtained. Likewise, HIV-1 integration sites tallied, respectively 69, 125 and 21 tracks in these categories. Sixty-two percent of the ENCODE tracks scored significant correlation coefficients of the same sign for LEDGF/p75 binding and HIV-1 integration sites (green squares; Figure 5), and only 1.8% scored significant correlation coefficients of the opposite sign. This analysis corroborates the proposed role of LEDGF/p75 in targeting of HIV-1 integration. An overview of all tracks can be found in the Supplementary Data (Supplementary Table S2). Cross-correlation enables comparison between

corresponding chromatin features, but also between shifted ones as revealed by cross-correlation curves (Supplementary Figure S3). As shown in Figure 6, LEDGF/p75 binding correlated positively with markers of euchromatin and transcriptional activation such as H3 and H4 acetylation, H3K4 monomethylation and RNA polymerase II binding, and negatively with determinants of heterochromatin such as H3K27 trimethylation. Remarkably, although HIV-1 integration also positively correlates with these markers of active transcription as published before (28), HIV-1 integration appeared to be scattered around the LEDGF/p75 peak (Figure 6).

Differential association of LEDGF/p75 and HIV-1 integration with ENCODE tracks

Interestingly, some ENCODE tracks such as Stat1 (signal transducer and activator of transcription), Hnf4a, Hnf3b and Usf1 chromatin binding correlated strongly with LEDGF/p75, but not with HIV-1 integration (Figure 7). After viral induction of interferon- γ , Stat1 is known to translocate from the cytoplasm to the nucleus, where it binds and activates target genes (48). Stat1 is strongly associated with LEDGF/p75 binding but not with HIV-1 integration (Figure 7, upper left panel). The same goes for the functionally related transcription factors, Hnf4a, Hnf3b and Usf1 (49), even though the coefficient of correlation of LEDGF/p75 with these factors is an order of magnitude lower than that between Hnf3b and Hnf4a or Usf1 (data not shown). These examples document for the first time that not all LEDGF/p75 complexes bound to chromatin are accessible for the HIV-1 pre-integration complex to support integration or that integration at these positions is restricted.

DISCUSSION

Research hints at a role of LEDGF/p75 in stress response [See (12) and references therein]. LEDGF/p75 expression is induced by oxidative stress and LEDGF/p75 itself is believed to activate stress-related genes by binding stress response and heat shock-related elements (4). In this work, we defined the LEDGF/p75 chromatin binding profile in the ENCODE region using DamID. DamID technology is extensively used in *Drosophila* genomics and is as powerful as ChIP-on-chip to determine chromatin binding profiles (50). While the resolution is the same as for ChIP-on-chip (1–2 kb), DamID has the advantage that neither antibody nor cross-linking is needed. Furthermore, expression of the Dam fusion protein at extremely low level prevents aspecific and saturating methylation levels, as well as disturbance of the normal cellular physiology.

Analysis of the resulting LEDGF/p75 binding islands revealed that 75% of LEDGF/p75 islands were located in TUs although the promoter region itself was disfavored. TUs are not enriched for 'gata' sites as compared to the whole ENCODE region ($P < 0.05$, Mann–Whitney U-test). Hence, enrichment of the LEDGF/p75 binding islands within TUs cannot be attributed to the 'gata' dependence of the DamID technology, but is indicative of

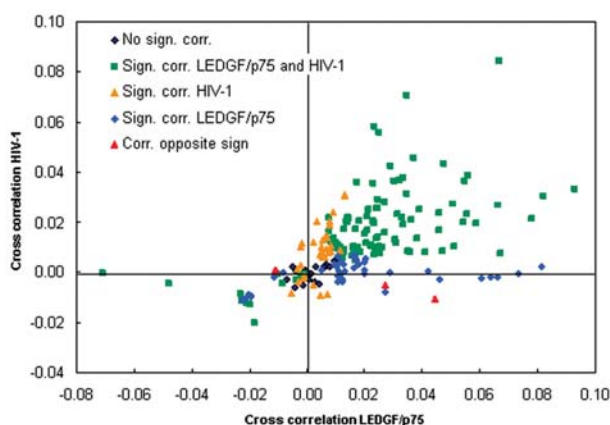


Figure 5. Cross-correlation of LEDGF/p75 islands and HIV-1 integration with ENCODE tracks. Scatter plot of the correlation of each of the 215 ENCODE tracks with LEDGF/p75 islands and HIV-1 integration sites. Significance was tested by calculating the 2.5 through 97.5 percentile interval of Pearson's correlation coefficients for 600 HIV-1 and LEDGF/p75 matched random controls with each ENCODE track of interest as described in the 'Materials and Methods' section. Most ENCODE tracks with a positive correlation with HIV-1 integration, positively correlate with LEDGF/p75 binding (in green). Black, no correlation with HIV-1 or LEDGF/p75 track; orange, positive or negative correlation with HIV-1 integration, no correlation with LEDGF/p75 binding; blue, positive or negative correlation with LEDGF/p75 binding, no correlation with HIV-1 integration; red, correlation of the HIV-1 and LEDGF/p75 track with the ENCODE track have the opposite sign.

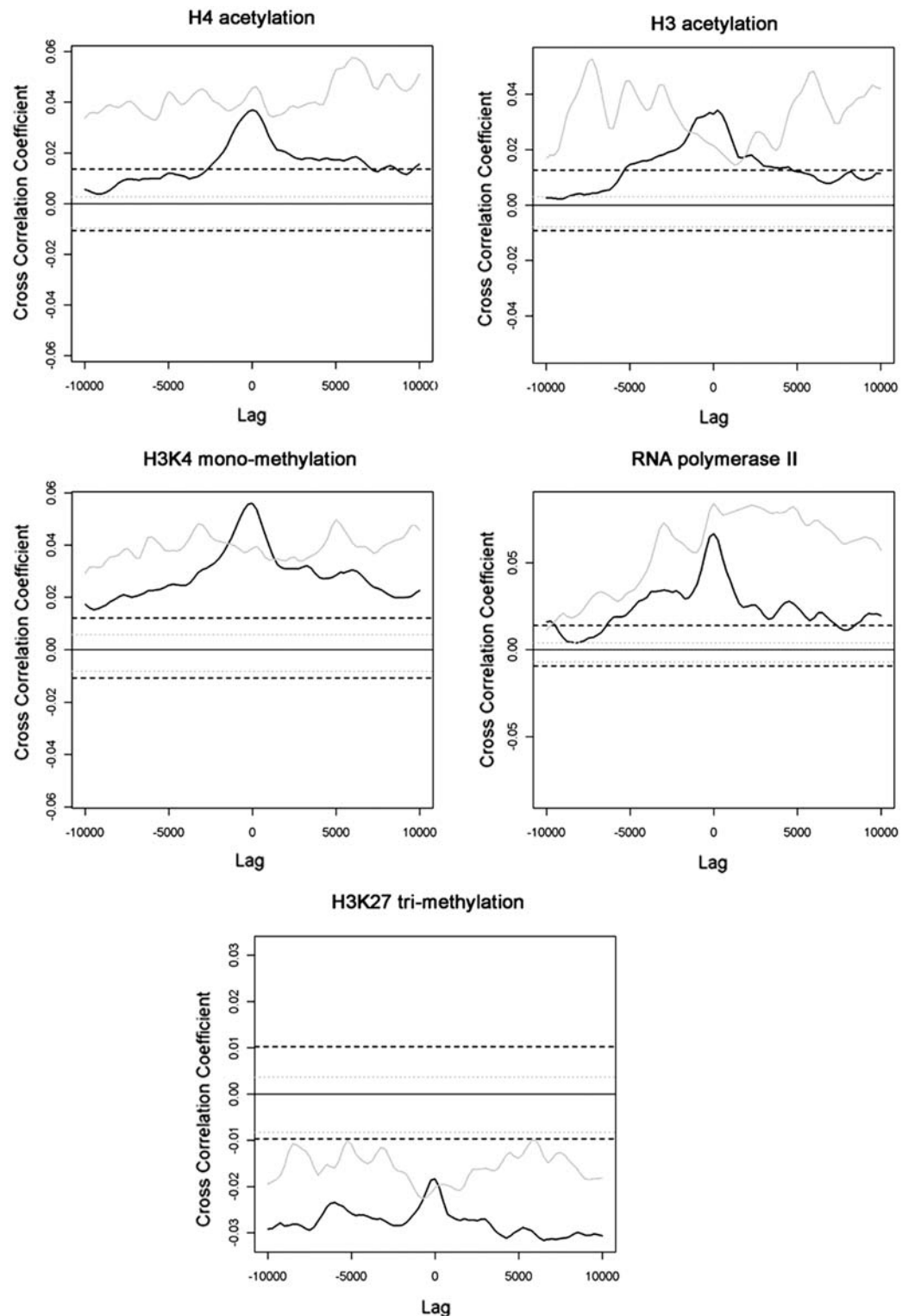


Figure 6. Shifted cross-correlation of LEDGF/p75 islands and HIV-1 integration with ENCODE tracks. The Pearson's correlation coefficient between LEDGF/p75 (black) or HIV-1 integration sites (gray) and the indicated ENCODE track was calculated for each lag. Lags indicate the distance the LEDGF/p75 or HIV-1 integration track was shifted as compared to the ENCODE track of interest to calculate the cross-correlation coefficient. Both for LEDGF/p75 islands and HIV-1 integration sites, 600 random tracks of the same size were generated. The correlation coefficient at lag 0 of the depicted ENCODE track with each random track was calculated. The 2.5 through 97.5 percentile interval fell between the dashed lines (black, LEDGF/p75; gray, HIV-1 integration sites). Histones 3 and 4 (H3 and H4) acetylation and Histone 3 Lysine 4 (H3K4) monomethylation are epigenetic markers of active transcription. Histone 3 Lysine 27 (H3K27) trimethylation is a marker of heterochromatin.

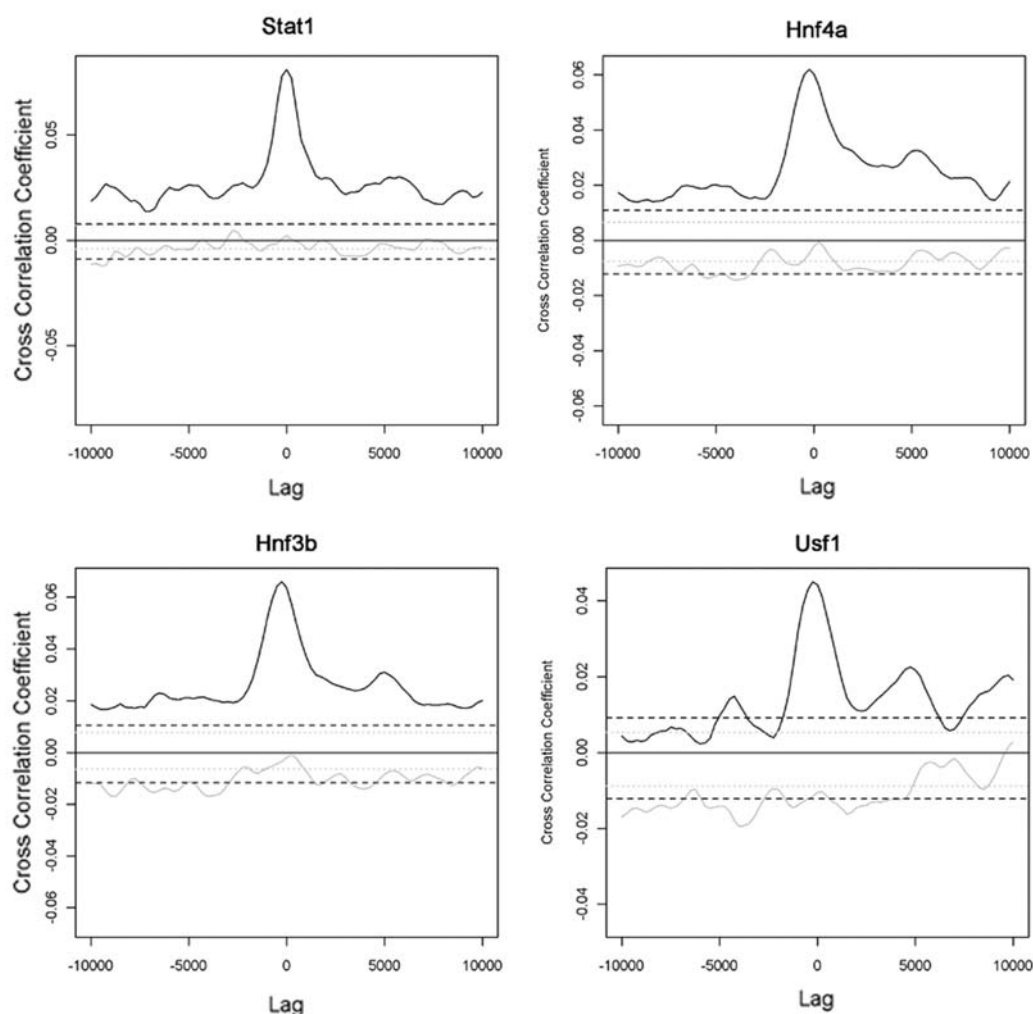


Figure 7. Divergent cross-correlation of LEDGF/p75 islands and HIV-1 integration with specific ENCODE tracks. For an explanation of the cross-correlation curves, see legend to Figure 6. The ENCODE tracks shown have a strong positive correlation with LEDGF/p75, but do not correlate with HIV-1 integration. Stat1, signal transducer and activator of transcription; Hnf3b, hepatocyte nuclear factor 3- β ; Hnf4a, hepatocyte nuclear factor 4- α ; Usf1, upstream transcription factor 1.

genuine LEDGF/p75 enrichment in these regions. LEDGF/p75 binds downstream of the TSS with a frequency decreasing towards the end of the TU. These results seem at odds with our current understanding of the cellular function of LEDGF/p75, since an activator of expression of stress response genes would be expected to be enriched at promoter sites. Though some studies indicated that LEDGF/p75 binds specific sequence elements associated with the promoter regions of stress response genes (4,6,31,32), others failed to detect sequence-specific LEDGF/p75 binding activity *in vitro* (14). It can however not be excluded that LEDGF/p75 binds stress response elements during stress conditions. Of note, the ENCODE region does not contain stress-responsive genes. A future DamID experiment with stress-induced cells could shed light on this issue. Nevertheless, at this stage our results demonstrate that LEDGF/p75 chromatin binding is not limited to stress response genes since it amounts to 28% of the ENCODE RefSeq genes. Although the proportion of LEDGF/p75-bound genes might be underestimated due

to the strict threshold used to analyze the microarray data, this is significantly less than the 37% for the corresponding control set, suggesting that LEDGF/p75 targets only a specific subset of genes. Though we have to take into account the relatively small amount of genes present in the ENCODE region, Gene Ontology analysis did not support any significant functional enrichment for certain gene product characteristics. Twenty-five percent of the LEDGF/p75 islands located outside of TUs. The mean score and size of this fraction is indistinguishable from that of all islands suggesting that these represent genuine LEDGF/p75 islands. In addition, the mean distance of this fraction to TUs does not differ from random (data not shown). The function of those islands remains to be investigated. At this stage, one can only speculate that these islands may have an important function in the establishment of latent lentiviral proviruses.

As expected and based on the known association between LEDGF/p75 binding and HIV-1 integration, the LEDGF/p75 chromatin interaction profile is reminiscent of that of HIV-1 integration (28). Much alike HIV-1

integration, LEDGF/p75 binding prefers the body of genes, disfavoring the promoter regions and correlating with transcriptional activity. Compared to 75% of the LEDGF/p75 islands, 86% of the ENCODE region-associated HIV integration sites were found in a TU and 30% of all integration sites were found in a window of 3.5 kb around the center of an LEDGF/p75 island, which amounts to a more than 3-fold enrichment over control integration sites. These data corroborate that LEDGF/p75 plays a role in HIV-1 targeting. This apparent window might however be influenced by the resolution and sensitivity of the DamID technology. Moreover, our DamID experiments were carried out in HeLaP4 cells, while the integration data set was derived from a Jurkat T cell line. Nevertheless, controlling our analysis for different expression profiles or chromosomal content between both cell lines did not significantly change the obtained results.

The correlation of the majority of the more than 200 studied ENCODE tracks with the LEDGF/p75 binding profile often mirrors that of HIV integration [this article and (28)]. Overall, LEDGF/p75 binding was associated with markers of active transcription like H3 and H4 acetylation, H3K4 monomethylation and RNA polymerase II binding, but correlated negatively with markers of heterochromatin. The cross-correlation curves with LEDGF/p75 binding sites revealed interesting patterns and some striking differences with HIV integration sites. In most cases, ENCODE tracks with a strong correlation with LEDGF/p75 chromatin binding also showed a high coefficient of correlation with HIV integration. While the correlation with LEDGF/p75 peaked over a relatively small window, that with HIV integration was more spread out around the LEDGF/p75 peak, in line with the window of enrichment of HIV-1 integration straddling LEDGF/p75 binding spots. These results indicate that the DamID resolution is high enough for comparison of LEDGF/p75 chromatin interaction with HIV integration and again suggests that HIV integrates in the wide neighborhood of LEDGF/p75 binding.

Interestingly, our data point out that not all chromatin bound LEDGF/p75 supports effective HIV integration. Indeed, the transcription factors Stat1, Hnf4a, Hnf3b and Usf1 correlated well with LEDGF/p75 binding but not with HIV-1 integration. The chromatin binding profile of this integration incompatible LEDGF/p75 fraction is indistinguishable from that of the complete LEDGF/p75 track (data not shown). HIV-1 integrase interacts with the integrase binding domain of LEDGF/p75, which is known to bind as well to other proteins like Jpo2, pogZ, MLL/men1 and Cdc7-ASK (7,15–18). It will be of interest to verify whether the chromatin binding of the LEDGF/p75 fraction that is incompatible for HIV integration correlates with the binding profile of one of these alternative partners. Competition in the binding with LEDGF/p75 may abrogate efficient integration.

In conclusion, the LEDGF/p75 chromatin binding profile corroborates the previously claimed association between LEDGF/p75 binding and HIV-1 integration. Still, other determinants seem to play a role since not all LEDGF/p75 sites support HIV integration, and

integration can occur at some distance from the actual LEDGF/p75 chromatin interaction spot. Moreover, our data challenge the current concept on the role of LEDGF/p75 in cell metabolism. It is clear that the function of LEDGF/p75 is not restricted to stress response. The more general chromatin interaction profile is compatible with a global role in transcriptional regulation. Of note, LEDGF/p75 was originally identified as a component of the general transcriptional machinery (1). In this regard, it will be of interest to analyze the expression and chromatin interaction profile of LEDGF/p75 in response to cellular stress.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

ACKNOWLEDGEMENTS

We thank Rudy van Eijsden of the KULeuven microarray facility for help with the microarray analysis, the KULeuven Cell Imaging Core for use of the confocal microscope, Prof. Peter Vermeersch, Center for Human Genetics, KULeuven for help with the cytogenetic analysis and Prof. Bas van Steensel, Amsterdam Cancer Institute, The Netherlands for providing the DamID constructs and the discussions on obtained results.

FUNDING

CellCoVir SBO grant (60813) of the agency for Innovation by Science and Technology (IWT); the Research Foundation Flanders (FWO) (grant G.0530.08); European Commission grant THINC (HEALTH-F3-2008-201032); Mathilde-Krim postdoctoral fellowship from Amfar (to J.D.R.). Funding for open access charge: THINC.

Conflict of interest statement. None declared.

REFERENCES

- Ge, H., Si, Y. and Roeder, R.G. (1998) Isolation of cDNAs encoding novel transcription coactivators p52 and p75 reveals an alternate regulatory mechanism of transcriptional activation. *EMBO J.*, **17**, 6723–6729.
- Sharma, P., Fatma, N., Kubo, E., Shinohara, T., Chylack, L.T. Jr and Singh, D.P. (2003) Lens epithelium-derived growth factor relieves transforming growth factor-beta1-induced transcription repression of heat shock proteins in human lens epithelial cells. *J. Biol. Chem.*, **278**, 20037–20046.
- Sharma, P., Singh, D.P., Fatma, N., Chylack, L.T. Jr and Shinohara, T. (2000) Activation of LEDGF gene by thermal and oxidative stresses. *Biochem. Biophys. Res. Commun.*, **276**, 1320–1324.
- Singh, D.P., Fatma, N., Kimura, A., Chylack, L.T. Jr and Shinohara, T. (2001) LEDGF binds to heat shock and stress-related element to activate the expression of stress-related genes. *Biochem. Biophys. Res. Commun.*, **283**, 943–955.
- Singh, D.P., Ohguro, N., Kikuchi, T., Sueno, T., Reddy, V.N., Yuge, K., Chylack, L.T. Jr, Shinohara, T., Fatma, N. and Kimura, A. (2000) Lens epithelium-derived growth factor: effects on growth and survival of lens epithelial cells, keratinocytes, and fibroblasts. *Biochem. Biophys. Res. Commun.*, **267**, 373–381.

6. Brown-Bryan, T.A., Leoh, L.S., Ganapathy, V., Pacheco, F.J., Mediavilla-Varela, M., Filippova, M., Linkhart, T.A., Gijbsers, R., Debyser, Z. and Casiano, C.A. (2008) Alternative splicing and caspase-mediated cleavage generate antagonistic variants of the stress oncoprotein LEDGF/p75. *Mol. Cancer Res.*, **6**, 1293–1307.
7. Yokoyama, A. and Cleary, M.L. (2008) Menin critically links MLL proteins with LEDGF on cancer-associated target genes. *Cancer Cell*, **14**, 36–46.
8. Ahuja, H.G., Hong, J., Aplan, P.D., Tcheurekdjian, L., Forman, S.J. and Slovak, M.L. (2000) t(9;11)(p22;p15) in acute myeloid leukemia results in a fusion between NUP98 and the gene encoding transcriptional coactivators p52 and p75-lens epithelium-derived growth factor (LEDGF). *Cancer Res.*, **60**, 6227–6229.
9. Grand, F.H., Koduru, P., Cross, N.C. and Allen, S.L. (2005) NUP98-LEDGF fusion and t(9;11) in transformed chronic myeloid leukemia. *Leuk. Res.*, **29**, 1469–1472.
10. Hussey, D.J., Moore, S., Nicola, M. and Dobrovic, A. (2001) Fusion of the NUP98 gene with the LEDGF/p52 gene defines a recurrent acute myeloid leukemia translocation. *BMC Genet.*, **2**, 20–24.
11. Morerio, C., Aquila, M., Rosanda, C., Rapella, A., Tassano, E., Micalizzi, C. and Panarello, C. (2005) t(9;11)(p22;p15) with NUP98-LEDGF fusion gene in pediatric acute myeloid leukemia. *Leuk. Res.*, **29**, 467–470.
12. Ganapathy, V., Daniels, T. and Casiano, C.A. (2003) LEDGF/p75: a novel nuclear autoantigen at the crossroads of cell survival and apoptosis. *Autoimmun. Rev.*, **2**, 290–297.
13. Llano, M., Vanegas, M., Hutchins, N., Thompson, D., Delgado, S. and Poeschla, E.M. (2006) Identification and characterization of the chromatin-binding domains of the HIV-1 integrase interactor LEDGF/p75. *J. Mol. Biol.*, **360**, 760–773.
14. Turlure, F., Maertens, G., Rahman, S., Cherepanov, P. and Engelman, A. (2006) A tripartite DNA-binding element, comprised of the nuclear localization signal and two AT-hook motifs, mediates the association of LEDGF/p75 with chromatin in vivo. *Nucleic Acids Res.*, **34**, 1663–1675.
15. Bartholomeeusen, K., Christ, F., Hendrix, J., Rain, J.C., Emiliani, S., Benarous, R., Debyser, Z., Gijbsers, R. and De Rijck, J. (2009) Lens epithelium-derived growth factor/p75 interacts with the transposase-derived DDE domain of Pogo. *J. Biol. Chem.*, **284**, 11467–11477.
16. Bartholomeeusen, K., De Rijck, J., Busschots, K., Desender, L., Gijbsers, R., Emiliani, S., Benarous, R., Debyser, Z. and Christ, F. (2007) Differential interaction of HIV-1 integrase and JPO2 with the C terminus of LEDGF/p75. *J. Mol. Biol.*, **372**, 407–421.
17. Hughes, S., Jenkins, V., Dar, M.J., Engelman, A. and Cherepanov, P. (2010) Transcriptional co-activator LEDGF interacts with Cdc7-activator of S-phase kinase (ASK) and stimulates its enzymatic activity. *J. Biol. Chem.*, **285**, 541–554.
18. Maertens, G.N., Cherepanov, P. and Engelman, A. (2006) Transcriptional co-activator p75 binds and tethers the Myc-interacting protein JPO2 to chromatin. *J. Cell Sci.*, **119**, 2563–2571.
19. Cherepanov, P., Devroe, E., Silver, P.A. and Engelman, A. (2004) Identification of an evolutionarily conserved domain in human lens epithelium-derived growth factor/transcriptional co-activator p75 (LEDGF/p75) that binds HIV-1 integrase. *J. Biol. Chem.*, **279**, 48883–48892.
20. Vanegas, M., Llano, M., Delgado, S., Thompson, D., Peretz, M. and Poeschla, E. (2005) Identification of the LEDGF/p75 HIV-1 integrase-interaction domain and NLS reveals NLS-independent chromatin tethering. *J. Cell Sci.*, **118**, 1733–1743.
21. Busschots, K., Vercammen, J., Emiliani, S., Benarous, R., Engelborghs, Y., Christ, F. and Debyser, Z. (2005) The interaction of LEDGF/p75 with integrase is lentivirus-specific and promotes DNA binding. *J. Biol. Chem.*, **280**, 17841–17847.
22. Cherepanov, P. (2007) LEDGF/p75 interacts with divergent lentiviral integrases and modulates their enzymatic activity in vitro. *Nucleic Acids Res.*, **35**, 113–124.
23. Cherepanov, P., Maertens, G., Proost, P., Devreese, B., Van Beeumen, J., Engelborghs, Y., De Clercq, E. and Debyser, Z. (2003) HIV-1 integrase forms stable tetramers and associates with LEDGF/p75 protein in human cells. *J. Biol. Chem.*, **278**, 372–381.
24. Llano, M., Vanegas, M., Fregoso, O., Saenz, D., Chung, S., Peretz, M. and Poeschla, E.M. (2004) LEDGF/p75 determines cellular trafficking of diverse lentiviral but not murine oncoretroviral integrase proteins and is a component of functional lentiviral preintegration complexes. *J. Virol.*, **78**, 9524–9537.
25. Ciuffi, A. and Bushman, F.D. (2006) Retroviral DNA integration: HIV and the role of LEDGF/p75. *Trends Genet.*, **22**, 388–395.
26. Schroder, A.R., Shinn, P., Chen, H., Berry, C., Ecker, J.R. and Bushman, F. (2002) HIV-1 integration in the human genome favors active genes and local hotspots. *Cell*, **110**, 521–529.
27. Shun, M.C., Raghavendra, N.K., Vandegraaff, N., Daigle, J.E., Hughes, S., Kellam, P., Cherepanov, P. and Engelman, A. (2007) LEDGF/p75 functions downstream from preintegration complex formation to effect gene-specific HIV-1 integration. *Genes Dev.*, **21**, 1767–1778.
28. Wang, G.P., Ciuffi, A., Leipzig, J., Berry, C.C. and Bushman, F.D. (2007) HIV integration site selection: analysis by massively parallel pyrosequencing reveals association with epigenetic modifications. *Genome Res.*, **17**, 1186–1194.
29. Ciuffi, A., Llano, M., Poeschla, E., Hoffmann, C., Leipzig, J., Shinn, P., Ecker, J.R. and Bushman, F. (2005) A role for LEDGF/p75 in targeting HIV DNA integration. *Nat. Med.*, **11**, 1287–1289.
30. Marshall, H.M., Ronen, K., Berry, C., Llano, M., Sutherland, H., Saenz, D., Bickmore, W., Poeschla, E. and Bushman, F.D. (2007) Role of PSIP1/LEDGF/p75 in lentiviral infectivity and integration targeting. *PLoS ONE*, **2**, e1340.
31. Fatma, N., Singh, D.P., Shinohara, T. and Chylack, L.T. Jr (2001) Transcriptional regulation of the antioxidant protein 2 gene, a thiol-specific antioxidant, by lens epithelium-derived growth factor to protect cells from oxidative stress. *J. Biol. Chem.*, **276**, 48899–48907.
32. Sapoznik, S., Cohen, B., Tzuman, Y., Meir, G., Ben-Dor, S., Harmelin, A. and Neeman, M. (2009) Gonadotropin-regulated lymphangiogenesis in ovarian cancer is mediated by LEDGF-induced expression of VEGF-C. *Cancer Res.*, **69**, 9306–9314.
33. Geraerts, M., Michiels, M., Baekelandt, V., Debyser, Z. and Gijbsers, R. (2005) Upscaling of lentiviral vector production by tangential flow filtration. *J. Gene Med.*, **7**, 1299–1310.
34. Gijbsers, R., Ronen, K., Vets, S., Malani, N., De Rijck, J., McNeely, M., Bushman, F.D. and Debyser, Z. (2010) LEDGF hybrids efficiently retarget lentiviral integration into heterochromatin. *Mol. Ther.*, **18**, 552–560.
35. De Rijck, J., Vandekerckhove, L., Gijbsers, R., Hombrouck, A., Hendrix, J., Vercammen, J., Engelborghs, Y., Christ, F. and Debyser, Z. (2006) Overexpression of the lens epithelium-derived growth factor/p75 integrase binding domain inhibits human immunodeficiency virus replication. *J. Virol.*, **80**, 11498–11509.
36. Vogel, M.J., Peric-Hupkes, D. and van Steensel, B. (2007) Detection of in vivo protein-DNA interactions using DamID in mammalian cells. *Nat. Protoc.*, **2**, 1467–1478.
37. Johnson, W.E., Li, W., Meyer, C.A., Gottardo, R., Carroll, J.S., Brown, M. and Liu, X.S. (2006) Model-based analysis of tiling-arrays for ChIP-chip. *Proc. Natl Acad. Sci. USA*, **103**, 12457–12462.
38. Li, C. and Hung Wong, W. (2001) Model-based analysis of oligonucleotide arrays: model validation, design issues and standard error application. *Genome Biol.*, **2**, RESEARCH0032.
39. van Steensel, B. and Henikoff, S. (2000) Identification of in vivo DNA targets of chromatin proteins using tethered dam methyltransferase. *Nat. Biotechnol.*, **18**, 424–428.
40. Cherepanov, P., Sun, Z.Y., Rahman, S., Maertens, G., Wagner, G. and Engelman, A. (2005) Solution structure of the HIV-1 integrase-binding domain in LEDGF/p75. *Nat. Struct. Mol. Biol.*, **12**, 526–532.
41. ENCODE consortium, The (2004) The ENCODE (ENCyclopedia Of DNA Elements) Project. *Science*, **306**, 636–640.
42. Beissbarth, T. and Speed, T.P. (2004) Gostat: find statistically overrepresented Gene Ontologies within a group of genes. *Bioinformatics*, **20**, 1464–1465.
43. Ng, P. and Keich, U. (2008) GIMSAN: a Gibbs motif finder with significance analysis. *Bioinformatics*, **24**, 2256–2257.

44. Ferris,A.L., Wu,X., Hughes,C.M., Stewart,C., Smith,S.J., Milne,T.A., Wang,G.G., Shun,M.C., Allis,C.D., Engelman,A. *et al.* (2010) Lens epithelium-derived growth factor fusion proteins redirect HIV-1 DNA integration. *Proc. Natl Acad. Sci. USA*, **107**, 3135–3140.
45. Silvers,R.M., Smith,J.A., Schowalter,M., Litwin,S., Liang,Z., Geary,K. and Daniel,R. (2010) Modification of integration site preferences of an HIV-1-based vector by expression of a novel synthetic protein. *Hum. Gene Ther.*, **21**, 337–349.
46. Berry,C., Hannenhalli,S., Leipzig,J. and Bushman,F.D. (2006) Selection of target sites for mobile DNA integration in the human genome. *PLoS Comput. Biol.*, **2**, e157.
47. Mitchell,R.S., Beitzel,B.F., Schroder,A.R., Shinn,P., Chen,H., Berry,C.C., Ecker,J.R. and Bushman,F.D. (2004) Retroviral DNA integration: ASLV, HIV, and MLV show distinct target site preferences. *PLoS Biol.*, **2**, E234.
48. Schindler,C., Fu,X.Y., Improtà,T., Aebersold,R. and Darnell,J.E. Jr (1992) Proteins of transcription factor ISGF-3: one gene encodes the 91-and 84-kDa ISGF-3 proteins that are activated by interferon alpha. *Proc. Natl Acad. Sci. USA*, **89**, 7836–7839.
49. Odom,D.T., Dowell,R.D., Jacobsen,E.S., Nekludova,L., Rolfe,P.A., Danford,T.W., Gifford,D.K., Fraenkel,E., Bell,G.I. and Young,R.A. (2006) Core transcriptional regulatory circuitry in human hepatocytes. *Mol. Syst. Biol.*, **2**, 2006.0017.
50. van Steensel,B. and Henikoff,S. (2003) Epigenomic profiling using microarrays. *Biotechniques*, **35**, 346–356.
51. Vogel,M.J., Guelen,L., de Wit,E., Peric-Hupkes,D., Loden,M., Talhout,W., Feenstra,M., Abbas,B., Classen,A.K. and van Steensel,B. (2006) Human heterochromatin proteins form large domains containing KRAB-ZNF genes. *Genome Res.*, **16**, 1493–1504.